ES DSWG Meeting

Baltimore MD
Oct 25-27, 2005

Tuesday October 25

Participants : Rich Ullman, Yonsook Enloe, Helen Conover, Allan Doyle, Ananth Rao, Ming-Hsiang Tsou, Jingli Yang, James Gallagher, Sam Bacharach, Chris Lynnes, Oscar Castenedas, Ross Swick, Ken Keiser, Don Faber-Langendoen, Gi-Kong Kim, Liping Di, Ken McDonald, Glenn Cunningham

-- Standards Breakout 1 – Community Presentations --

Chris Lynnes – standards for reuse and interoperability at GES DAAC

- Assumption that a standard will support interoperability, **implement those that promote reusability within GES DAAC** as well (requirements, design, implementation). Difficult to gauge utility and usability of interoperability. Easier to determine whether standard can be reused.
- PDR/PAN – operational, high performance, used locally w/in DAAC; but whole files only
- OPeNDAP – implemented for data serving, no cataloging yet; subsetting very handy; some concern about performance overhead
- OpenGIS – using WMS only now (Minnesota Map Server, Synergy Standalone OGC Server). Takes a fair amount of tailoring and WMS is not friendly to time and depth/height dimensions. Catalog is still evolving.

Li Ping, Sam: WCS will be better (parameters, time, etc.)
Chris: but still many implementation decisions to make….

- ECHO file level metadata spec – reused for S4PA (disk-based data mgmt and archive system)
- GCMD collection level metadata spec – nice web forms, etc., convertible to XML, FGDC; but cede control of collection level metadata to another organization
- Web Services – evolve to web services plug and play components at DAAC: Mirador search; Giovanni data analysis; WHOM order engine; FCP subset service. Need to standardize context (dataset reference, parameters, space and time).
- Grid – considering to link distributed systems – but tricky…

Gallagher: Another reason for standards is quality, leveraging past knowledge
Rich: PDR/PAN as potential RFC???
Chris: Stability of standard is important, want backward compatibility. Implementation cost is critical.

Liping Di – OGC WCS

- **Contrast with DAP: WCS uses map coordinates, DAP uses data array coordinates.** WCS is one level more abstract than DAP2. Advantage: allows client to define the retrieved data by specifying real world location and data resolution. Multiple source data integration will become automatic.
- GetCapabilities (list of all coverages offered), DescribeCoverage (details on any given coverage) for XML metadata
- Data must be in one of five well-defined formats: HDF-EOS, GeoTIFF, DTED, NITF, GML. OGC GALEON working on netCDF.

Lynnes: are clients really implementing z dimension?
Liping: GALEON working with 5D data (x, y, z ,t, data parameter)
Sam: Space Time Toolkit does

Ken Keiser – ESML

- Motivation – provides metadata not readily available elsewhere, coming into wider use
- Prototype using ESML to describe OGC Coverages
- Benefits of ESML : software access to heterogeneous data formats. Software insulation to format changes – change the description no the code. Single description for file groups that are syntactically and semantically similar.
- Have implemented OPeNDAP/ESML server to increase the number of data formats supported by OPeNDAP.

Lynnes: Perl support?
Elena: Need help from HDF team for HDF-4 support
Lynnes: How to identify geographic dimensions – use specific conventions for that?
Ming: compete with FGDC? Consider trying to add to FGDC.
Ken: complementary, used for different reasons
Oscar: vision?
Ken: One option is registry of descriptions, also could be generated by users as needed.
Chris: could also put ESML behind WMS/WCS, similar to OPeNDAP implementation

Don Faber-Langendorn – FGDC 1997 US National Vegetation Classification standard – classification standard and/or attribute standard. Foster accuracy, consistency, and clarity in the structure labeling, definition, and application of a systematic vegetation taxonomy for the US. A list of guiding principles for what constitutes a good set of classification units. Community of reviewers is US agencies, Canada, some Latin American countries. Users: anyone collecting vegetation plot data (students, academics, consultants, agencies). Federal land management agencies are users. How much of the user community is NASA? Is this area an area for science research or infrastructure building?

Rich: How would NASA use this standard? Provide maps using this system as a separate product? Some disagreement on how to encode this data.
Don: Important to agree on attributes so anyone can implement. Also important to provide perspective on what classification might be across jurisdictions. Which attributes do you want to store?

Sam: efforts to coordinate with international partners provide flexibility. Not definitive answer, but FGDC standards reviewed every 5 years.
Ming: what are competitive standards?
Don: Nat'l land cover database

Oscar Castenedas – WebGIS technologies at TRFIC

- Basic question – are any of these potential standards?
- Search system relating books in library to specific scenes (Z39.50)
- Ready to add data server sites to search via web services or Z39.50.
- Can provide client software for others to use.

Allan: Consider **tech note** documenting design pattern. Consider publishing Landsat "convenience services" to translate between path-row and geographic coordinates
Lynnes: similar services for MODIS swaths would be useful

Ross Swick – Backtrack (see paper and brochure)

- Best method for doing spatial search of orbital swath data
- "Swath data is a weird shape – if you think the earth is flat…" How do you describe an orbit so that it's easy to see that it's an orbit and easy to search on? If the orbit described as a polygon and the polygon overlaps, then it's difficult for many database search engines to handle – could have large performance hit.
- Backtrack uses simplifying assumptions that apply very well to ES satellites, such as circular orbits. Assumes that the orbit is circular, then easier to describe.
- **The Backtrack RFC could be the Best Practice for searching earth science orbital swath data.**

Next steps: Tech note documenting a best practice – "request for use"
Software available? OrbitClass available in Spheres package
How should Tech Note look?

Wednesday Oct 26

-- Joint session – Standards / Reuse / Tech Infusion --

Discussion: Use Capabilities Vision and Roadmap developed by tech infusion group to plan joint or coordinated activities. ***Consider focusing on one area (maybe web services) and try to coordinate standards or conventions and reusable software that would facilitate wider infusion.***
Elaine: This mirrors V0 IMS development.

Discussion: Concern about pushing emerging practices through standards process rather than adopting current practice. But maybe use Tech Note to document conventions encouraging people to use common interfaces, etc. Who would write these Notes? Maybe TIWG would ask tech developer to do it.
Rich: We'd be asking people to write up things not already in common use, taking a risk. What's their incentive to make that effort?

Sam: Provide a place (on standards web site) for people to post suggested protocols or technologies as community resource. If these become widely used, then push through stds process.
Allan: Consider this an area for brainstorming. Once community distills this idea, then submit a tech note. Want high quality tech notes (best practices, etc.), not just a bunch of half-baked ideas.
Karen – three areas of common interest:

1. TIWG activity – emerging web services standards: tech note encouraging people to look at SOAP may be valuable
2. SPG activity – AURA metadata standard extending to atmos comp community: TIWG look at how this is evolving to see if it affects web services in any way
3. Potential community activity – content std for XML tag names: need forum to encourage exploration. Look for implementation projects, like GCMD, ECHO, Rob's ontology ACCESS project.

Other points:

- **R. Wolfe**: Timetable and/or goals for coming year? More formal coordination?
- **Elaine**: Create a virtual group from the three, to focus on one area and coordinate activities. Rest of groups can continue as planned.
- **Discussion**: Focus is critical (resources, achievable goals, etc.) Maybe pick a technology (web services) as applied to a specific science area.
- **Systems offered as test cases:**
    - Elaine – SciFlow, using web services to locate and co-register data.
    - Vic – digital atlas for intuitive search, would like to incorporate standards.

Actions:

- Document rationale for general case of coordination among groups – chairs of wgs
- Form focus group on web services among these three (and rest of community).

-- Standards Breakout 2 – HDF Discussion --

Participants: Rich Ullman, Larry Klein, Sam Bacharach, Al Fleig, Ananth Rao, Ming-Hsiang Tsou, Elena Pourmal, Allan Doyle, Siri Jodha Khalsa, Jingli Yang, Gi-Kong Kim, James Gallagher, Ken McDonald, Helen Conover, Yonsook Enloe, Peter L, Liping Di, Glenn Cunningham

Al Fleig on Aura Profile (Atmospheric Chemistry Data) for HDF-EOS 5 Guidelines
Aura experience demonstrates that it is developers at diverse locations creating multiple products from several instruments to agree on important common data formats. Guidelines are composed of (1) a set of guidelines explicitly defining format of similar data products from multiple instruments and (2) a set of naming conventions, units, metadata definitions (via HDF/HDF-EOS attributes) and data organization conventions. The goals of the standardization is to allow easier sharing and use of data and reduce development effort and support reuse by enabling g application of software to multiple instruments and products. It was difficult to get standard reviewed. Finally said – if you don't review it within x months, then it means you agree to it. Still didn't get much review. After 6 months, when standard was finalized, then project leader said, everyone

must conform. The stick was – the project leader owned the funding and owned the machines for the data processing. Any processing software that didn't conform with the standard, the project leader would not run the software. This ensured compliance. But within the wider community, we won't have this stick. Not sure what will happen then. Lessons Learned : Communicate early, before individual team's decision on data files have been made. Exchange data sets early on to assure common understanding of the standard. Include software engineers and scientists in the discussion group. Group leader must be firm, no issues left unresolved.

- Resources required to develop and maintain a living standard
- Resources required to maintain tools for use with data following the profile
- Considering extending use of this profile to atmospheric chemistry community
- Considering making this a content standard rather than an HDF profile
    - Field names, units, dimensions, data types

Issue: Should the Aura guidelines be tied to the HDF or be separated from it? All NASA remote sensing missions except one (GLAS) are using HDF. HDF is very real within the NASA community and not going to disappear. But may be useful to make the general case as well as the specific case. Write for the general case (not tied to HDF) and then document the profile for HDF and say, here's the HDF implementation of this general profile. Limited to Aura or extend this to the atmospheric chemistry community.
Discussion – two independent implementations: write Aura profile as a tech note, encourage other groups, probably the atmospheric chemistry community, to use it (independent implementations), then take to standards process.
Question: Best practices tech note to document process or to document profile?
Sam: FGDC producing set of framework layers for geo data, which may overlap with this effort.
Discussion – how stable is the profile? Level 2 "mostly done", Level 3 under consideration. Extended every time a new product is defined. Not truly done until last reprocessing. RFC can document practice and rationale, reference location of current document. For a Tech Note that's OK. Talk about in tomorrow's process discussion. **Near term activity is to get the Tech Note in.**

<u>Elena – HDF5 proposal for standardization</u>

- Vision for HDF – become PDF of scientific data

No official standard for general scientific binary data formats exists. 300 projects worldwide including NASA use HDF5. HDF5 standardization will validate HDF5 to vendors, government agencies, and other organizations.
Discussion – what to submit as RFC: Submitting HDF5v1.6 to ANSI, ISO already as international standard for scientific data exchange. Would submit same version to SPG. Standardize data model (20p), file format (50p), APIs (300p). Consider standardizing them separately. Most interested in seeing HDF5 as ANSI or ISO standard, specific profiles endorsed by SPG. Maybe SPG make recommendation that NASA help push HDF5 through ANSI or ISO standards process.
Discussion – HDF4: Where does HDF4 fit in? Should standardize HDF4? SPG is trying to facilitate interoperability. Should we be forward looking or worry about

heritage data? Is it appropriate to endorse HDF4 for wider use? Or just look forward to HDF5? A recent NASA mission, cloudsat, decided to use HDF4. But most newer missions expected to use HDF5 because of the better performance. May be appropriate to do HDF4 RFC where the motivation is to explain the current products in HDF4 but is not recommending for wider use. That way, there will be documentation to access HDF4 data – NASA has a lot of data in HDF4. Should document file format or the APIs?

Discussion – HDF-EOS as profile of HDF: Should HDF-EOS be endorsed by SPG? On top of HDF4 or HDF5? If SPG is looking to future, no need to recommend wider use of HDF4. However, it is important to document spec so others can implement if needed. What about long term support for HDF-EOS?
Discussion – what is a standard: Need to document the spec, or just recommend everyone use a particular package? That would be a tech note best practice.
Discussion – timing: Do we need to endorse (or wait for ISO to standardize) HDF before HDF-EOS before Aura profile?
Conclusions:

1. HDF Group will submit entire spec to SPG, no expectation for result (may be too big/complex for us to review), before going to ANSI and ISO.
2. Larry Klein will submit RFCs for HDF-EOS

Thursday Oct 27, 2005

■ *Standards Breakout 3: Process Discussion –*

Participants: Rich Ullman, Allan Doyle, James Gallagher, Ananth Rao, Glenn Cunningham, Siri Jodha Khalsa, Helen Conover, Jingli Yang, Yonsook Enloe, Ming-Hsiang Tsou

Modifications of RFCs

- Editorial changes – how long can we change them? Up until HQ approves, then after that we do errata? Editor approval only – changes that don't require updates to existing code.
- Technical changes – Changes that require coding change to existing software – new RFC
- Evolution of standard – additional features (backward compatible) – capture reviewed version and reference source for new versions

Endorse a standard or protocol (e.g., WMS) in general, while acknowledging we are reviewing a specific version. Users are encouraged to use the latest.

Are we a document repository or catalog?

- Problem of maintaining documents – ***keep a copy of any version we have reviewed and endorsed***
- Problem of maintaining links – ***provide a link to technology developer's web site for later versions.*** SPG website should run a dead link test every day.

James: valuable to OPeNDAP to host spec on NASA web site, because more authoritative.

After RFC is approved, further editorial changes captured in errata. After errata stack up (judgment call by editor – not necessarily original editor, but responsible person appointed by SPG), issue new version incorporating changes with internal review by SPG.

Helen will make sure process document reflects these conclusions.

Profiles of existing standards – don't have to review or endorse the original standard first. If profile is written such that the user needs a copy of the base standard also to implement the profile, then the SPG will keep a physical copy of the profile plus a physical copy of the base standard on the SPG website. If only the profile document is needed to implement the profile, then the SPG will keep a physical copy of only the profile document and provide a reference to the base standard document.

Need a checklist for editors – be sure to include:
- Expected stability of spec

<u>Standards maintained by other groups</u>
- ***Appropriate to review standards already maintained by other groups, because we're considering NASA DS use of the standard.***
- ***Can review a specific profile of a given standard without reviewing the parent standard.*** Profile must reference base standard. SPG should keep a copy of the base standard if possible, and provide link to current authoritative version.

<u>Community size</u>
- May be self limiting
  - interest in submitting RFC should relate to community size
  - requirement for two or more implementations (organizations behind it)
  - significant operational experience may not be there
- Small community may point to tech note – documenting and publishing can still have value
- This is what the initial review is for

<u>Outreach and tactics</u>
- No one has time to write RFC. May be valuable to NASA to have this documented. Funding?
  - Hire tech writer to work with technology developer? But this can show favoritism to certain groups.

    - Fellowship that interested groups can apply for
    - Include documenting specs in ROSES solicitations

Policy recommendations

- Language for future NRAs
    - Use of community standards
    - ***Documenting community practice as RFCs***
- Steering committee at NASA HQ level to handle policy implications and make policy recommendations.
    - One steering committee for all ES DSWGs
    - Composed of HQ decisions makers (Martha and other program managers)
        - Raise awareness at HQ of data systems value and issues
    - Recommendations to this group should come from working group chairs (not through Kathy only)
    - Suggest regular meetings between working group chairs and steering committee
    - One role of steering committee is to make policy recommendations for NASA programs (e.g., language for NRAs)
    - Another role is to provide guidance and direction to ES DSWGs – ***how we can be more effective and more useful to ES program?***
    - Some concern about this recommendation
        - Not to detract from Kathy and Martha's current work
        - Is the current organization working?
        - Maybe steering committee is not the right word for this concept
        - Maybe inviting program managers to these meetings or telecons is enough
    - ***Conclusion***: think about this for a while, discuss with other WGs.